DØ New Phenomena Analysis Approach

Greg Landsberg
February 13, 2003

Background

- New Phenomena group is the leader in terms of the number and the variety of the analyses
  - 7 Analyses for ICHEP 2002
  - 11 Analyses under the EB review for Winter 2003
  - Pretty much all the objects: jets, e, μ, ME_T, τ (!), and b's (no top yet!)
- A lot of experience accumulated with various data formats for the past 1.5 years: from root-tuples to TMBTrees
- This talk reflects our group experience, our praises and our gripes

What's Required for a Good Analysis?

- No good analysis can be done in a single iteration: multiple data reprocessing is always required

- An ability to run the analysis code on the highest data tier in real time is imperative

- An ability to run standard algorithms and corrections at some stage of the analysis is required

- An ability to trace luminosity information back to RAW data tier is a must

- Platform-independence of the algorithms and high-level data formats is desirable

New Phenomena Group Data Model

- We believe that on the long run the TMBTree format fits our analysis model
- It is reflected in the fact that all 11 analyses are using TMBTree tier
- It's essential for a successful analysis to use multi-tier data model in order to be able to go significantly far back without reprocessing RAW tier
- We achieve this via flexible skimming and small analysis samples
- Disk space is cheap; your time is not!
  - Don't try to save space by dropping chunks – you might regret later
  - Buy disk and fill it; when it's full buy some more or think how to deal with the data which is there
- In the case of the NP group, we raised $9K and bought 2.6 TB designated file server: insamity-clued0 with 2x8x180GB RAID arrays and a Gigabit Ethernet

New Phenomena TMB Streams

- TMB is the basic format for our offline skimming

1em	1emtrk	1em1mu	1mutrk
• 1 EM |ID| = 10, 11 pT > 20GeV Iso < 0.15 HM < 40 **4.4%**	• 1 EM |ID| = 10, 11 pT > 12GeV Iso < 0.15 • 1 Track pT > 7 GeV • dØ < 0.2 **1.3%**	• 1MuonCandidate Medium • 1 EM |ID| = 10, 11 pT > 5 GeV **2.0%**	• 1MuonCandidate Medium • 1 Track pT > 6 GeV • dØ < 0.2 **7.1%**

2em	2mu	1em2jets
• 2 EM |ID| = 10, 11 pT > 7GeV **1.9%**	• 2 MuonCandidate Medium **1.0%**	• 1 EM |ID| = 10, 11 pT > 15GeV Iso < 0.15 HM < 40 • 2jets jetName = JCCA pT > 15 GeV emETfraction <= 0.95 chETfraction <= 0.4 hotcellratio <= 10 n90 >= 1 **2.0%** • Trigger EM15_2JT15

Analysis Samples

- Every analysis starts with the TMB streams
- TMB Streams are converted in the TMBTrees, with the latest version (typically 2-3 days per stream)
- Typically, a preselection code is run on a particular set of the TMBTrees, which allows for a lot of flexibility (can rerun the entire preselection code in less than a day)
- Currently handicapped by the root bug that does not allow to save a subset of the TMBTree without destroying the cross-reference structure (TObjectArray bug, allegedly fixed in the 3.03.09 root release, to be confirmed)
- To work around this problem, a few different approaches were used:
  - Create a list of events on the fist pass, and then build a tree of these events only, which can be saved further (Alex)
  - Use old root-tuple format instead of TMBTrees (Serban)
  - Run all the corrections/ID routines on the TMBTree and output a very small analysis-specific root-tuple (Yuri)
  - Run all the corrections/ID routines on the TMBTree and output a roottuple, which contains most of the TMBTrees variables, with objects being identified by standard algorithms and references replaced with numerical cross-references (Greg) – this r-tuple can be used by any analysis

Conclusion

- We are generally satisfied with the TMBTree as the basic analysis platform
- We welcome the ID group efforts to make tools that work directly on the TMBTrees (e.g., EMcandidate, Mucandidate, JES)
- More tools of this kind are needed (e.g. ME_T corrections in the rings, etc.)
- Switching from one data format to another is a *major* time sink – please do not do it again!
- We would like the TMBTree format to be debugged and expanded, but we beg that it stays backward compatible and is used as our final attempt of the analysis data tier